1. The Standard Normal Distribution

Used to approximate or describe histograms of many (but not every) types of data. Properties are:

a. Symmetric, bell-shaped, the "bell curve", see page 86-87 of your textbook.

b. Mean 0, SD 1

c. The median is where 50% (half) of the observations are on either side. In this distribution, the mean is equal to the median. The values on the horizontal axis are called "Z SCORES" or "STANDARD UNITS". Values of Z above the average are positive, values of Z below the average are negative.

d. Area under the curve is equal to 100% when expressed as a percentage. The shaded area under the curve represents the percentages of the observations in your data to the left of given values of Z.

e. 68%-95%-99.7% rule (see p. 87) About 68% fall within plus or minus 1 SD of the mean About 95% fall within plus or minus 2 SD of the mean Nearly 100% (99.7%) fall within plus or minus 3 SD

f. The curve never crosses the horizontal axis, it gets very close at the extremes though. It extends to negative and positive infinity.

2. Standard (Deviation) Units

A score z is in STANDARD UNITS if tells how many SD's the original score is above or below the average. For example, if z=1.3, then the original score was 1.3 SD's above average; if z = -0.55, then the original score was 0.55 SD's BELOW average. The formula for converting data from original units to Z scores is:

$$\mathrm{z} = \frac{(\text{value of interest} - \text{average of all the values})}{\text{standard deviation of all the values}} = \frac{y-\mu}{\sigma}$$

you can call this a "normal calculation"

Note what the formula does, it subtracts out the population mean (mu) from a value and this has the effect of shifting the distribution in such a way that the a value that is equal to the population mean will be zero (see page 84). Dividing by sigma (the population standard deviation) will rescale the distribution (p. 85) in such a way that a value that is one standard deviation above the mean will take the value of 1, a value that is two standard deviations above the mean will take the value of 2 and so forth.

3. Changing Scales (pp.84-85)

Definition: changing the original values in a list of numbers by adding, subtracting, multiplying or dividing each value by a constant. All a Z score really does is change the scale of an existing set of values by subtracting the mean from a value and then dividing by the standard deviation. What effect does this have? It affects both "center" and "spread".

The mean will be changed (added to, subtracted from, multiplied or divided by the constants we used)

The standard deviation will be changed (multiplied by or divided by the constants we used)

- Essentially, this is how the standard normal curve works. It forces every normally distributed variable to have a mean of zero and a standard deviation of 1.
- The values of the mean, median, the minimum, maximum, percentiles, are affected by both the addition (or subtraction) of a constant and multiplication (or division) by a constant.

· The values of the measures of spread – that is, the IQR and the standard deviation, and the range, are only affected by multiplication (or division) by a constant, the IQR and standard deviation are not affected by the addition (or subtraction) of a constant.

4. Examples of the use of Standard Units

Law School Admissions Test scores are normally distributed with a mean of 150 and standard deviation of 10. It's range is 120-180. When last reported, the typical law student at Yale (the #1 law school) had an LSAT score of 171. We could express that in Z scores to give us a sense of "how high"

$$z = \frac{(171 - 150)}{10} = \frac{21}{10} = 2.10$$

or the typical Yale law student has a Z score of 2.10. This student is 2.10 standard deviations above average in their LSAT score. He or she scored higher than or 98.21% of all LSAT test takers. See Appendix E page A-83, look up 2.1, go to the first column and focus on the 98.21. The 96.43 is the shaded area or the total area (percentage) to the left of value 2.10

5. Converting Standard Units back to original values

Idea: suppose you are told that the minimum LSAT necessary for making the "first cut" at the UCLA school of law is Z=-1.15, what is the actual score? -1.15 = (value of interest - 150)/10 = about 138.5 or 139 on the LSAT.

6. Why bother with Standard Units?

Standard Units allow quick comparisons across variables with different units of measure. For example, suppose all the test scores in a class are normally distributed. The first test was worth 45 points, the mean was 33, the standard deviation was 5. A student received a 40 on that test and was told it was an A-, her Z score was

$$Z = (40 - 33)/5 = 1.40$$

The next test had a mean of 58 and standard deviation of 8. If she were to do as well on the second test as she did on the first (that is get an A-), what would her new score need to be?

$$Z = 1.40 = (\text{new score} - 58)/8$$

or she would need to score a 69.2 or something around a 69 to 70. The thing to remember is this: the second test has more points, a different mean, and a different standard deviation – it's different, but if we convert the "raw" scores to standard scores (Z), comparisons are easy. In a way, it lets us compare apples and oranges.

Try this, suppose you took the LSAT but decided not to go to Law School and instead applied to Graduate school but didn't take the GRE (Graduate Record Examination). Your LSAT score was 169, if you were to do as well on the GRE as you did on the LSAT, what would your GRE score be? Suppose GRE scores are normally distributed with a mean of 1000 and a standard deviation of 200.

The first thing you would need to do is convert your LSAT score in a Z score, so that's a

$$Z = (169 - 150)/10 = 1.90$$

And then translate the Z=1.90 into a GRE score or:

$$Z = 1.90 = (\text{new score} - 1000)/200$$

A 1380. Not bad.

7. Assessing Normality

A. Common sense: if the normal curve implies nonsense results (for example, that people have negative incomes, or that some women have a negative number of children), the normal curve doesn't apply and using the normal curve will give the wrong answer.

B. Construct a histogram: if the data look like a normal curve, the normal curve probably applies; otherwise, it does not.

C. Do the data fall in a 68-95-99.7% pattern? If yes, normality is probably being met.

D. NEVER ASSUME THAT A VARIABLE IS NORMAL.